Computer Science Department

TECHNICAL REPORT

ON THE DEPTH OF A RANDOM GRAPH

by

John H. Reif* and Paul Spirakis

July 1983

Technical Report No. 82

CSD TR-82
c.2

NEW YORK UNIVERSITY

Department of Computer Science
Courant Institute of Mathematical Sciences
251 MERCER STREET, NEW YORK, N.Y. 10012

ON THE DEPTH OF A RANDOM GRAPH

by

John H. Reif* and Paul Spirakis

July 1983

Technical Report No. 82

*Aiken Computation Laboratory, Harvard University.

This work was supported in part by the National Science Foundation Grant NSF-MCS 8300630.

SUMMARY

This paper investigates the probability distribution of the depth of a random graph $G_{n,p}$, or a random digraph $D_{n,p}$. We show the average depth of a random graph or digraph to be $O(\log n)$, for all (but a small range) of values of p. We furthermore show that the probability that the depth is $O(\log n)$ tends to 1 as n tends to infinity for all (but a small range) of values of p. We also prove that for $p \geqslant \frac{c}{\sqrt[3]{n}}$ (c a particular constant > 1) the depth d of $G_{n,p}$ or $D_{n,p}$ is less than or equal to 3 with probability tending to 1 as n tends to ∞. Although the result that the average depth of $G_{n,p}$ is $O(\log n)$ can be deduced from results of Erdös and Renyi for several values of p, the result for sparse graphs $\left(p = \theta\left(\frac{1}{n}\right)\right)$ and the result for very dense graphs $\left(p \geqslant \frac{c}{\sqrt[3]{n}}\right)$ are new contributions. The above results are useful in studies of the expected performance of parallel graph algorithms.

c 2

1. INTRODUCTION

DEFINITION: Let $G = (V,E)$ be a (directed or undirected) graph and let $d(v,w)$ for $v,w \in V$ be the length of the shortest path from v to w if any, $-\infty$ otherwise. Then the depth, $d(G)$, of the graph G is defined to be the $\max_{v,w \in V} \{d(v,w),1\}$. Note that for undirected graphs G with nontrivial connected components, $d(G)$ is the maximum of the diameters of the connected components of G.

The random graph $G_{n,p}$ is defined here as in [Erdös, Renyi, 60] and the random digraph $D_{n,p}$ as in [Angluin, Valiant, 79].

DEFINITION: For $0 \leqslant p \leqslant 1$ and $n \geqslant 0$ let $G_{n,p}$ ($D_{n,p}$) be a random variable whose values are graphs (digraphs) on the vertex set $\{1,2,..,n\}$. If $e = \{u,v\}$ (resp. $e = (u,v)$) and $u,v \in \{1,2,..,n\}$, $u \neq v$, then Prob {e is an edge} = p and these probabilities are independent for different e. Note that we refer here to simple linear graphs.

In the paper we investigate the probability distribution of the depth of a random graph $G_{n,p}$ or a random digraph $D_{n,p}$. The mean value $\bar{d}$ (and bounds on the probability distribution) of the depth are stated. We prove the following theorems:

THEOREM 1.1: There is a constant $c_0 > 2$ such that for any probability p in the range $[0,1] - \left[\frac{c_0}{n}, \frac{2c_0}{n} - \left(\frac{c_0}{n}\right)^2\right]$ the graph $G_{n,p}$ has average depth $\bar{d} = O(\log n)$. Furthermore, prob $\{d = O(\log n)\}$ tends to 1 as n tends to ∞.

THEOREM 1.2: There is a constant $c_0 > 2$ such that for any probability p in the range $[0,1] - \left[1 - \sqrt{1 - \frac{c_0}{n}}, \frac{c_0}{n}\right]$, the digraph $D_{n,p}$ has average depth $\bar{d} = O(\log n)$. Furthermore, prob $\{d = O(\log n)\}$ tends to 1 as n tends to ∞.

We also prove that for $p \geq \frac{c}{\sqrt[3]{n}}$ (c a particular constant > 1) the depth d of $G_{n,p}$ (or $D_{n,p}$) is less than or equal to 3 with probability $\to 1$ as $n \to \infty$. Although the result $\bar{d} = O(\log n)$ can be deduced from results of [Erdös, Renyi, 59] for several values of p (see text), the result for sparse graphs with edge probability $p = \theta\left(\frac{1}{n}\right)$ and the result for very dense graphs $\left(p \geq \frac{c}{\sqrt[3]{n}}\right)$ are new contributions.

2. PROOFS OF THEOREMS 1.1 AND 1.2

We prove here Theorems 1.1 and 1.2.

LEMMA 2.0: Theorem 1.1 implies Theorem 1.2.

PROOF: Let D be any instance of $D_{n,p}$. The depth of D cannot exceed the depth of the underlying undirected version G of D (which is produced from D by removing all edge directions). However, G is an instance of $G_{n,p}$, with $p' = 2p - p^2$. Theorem 1.1 holds for $G_{n,p'}$, for p' in the range

$$R = [0,1] - [\frac{c_0}{n}, \frac{2c_0}{n} - (\frac{c_0}{n})^2]$$

where c_0 is a fixed constant (to be determined in the proof of Theorem 1.1). Hence, Theorem 1.2 holds for $D_{n,p}$ for p determined by $p' = 2p - p^2 \in R$. □

2.1. Proof of Theorem 1.1 for $p = \theta(\frac{1}{n})$.

We prove here Theorem 1.1 for undirected graphs. This provides an upper bound on both the average depth and the depth distribution of $G_{n,p}$ for all $p = \theta(\frac{1}{n})$ except those p in $[\frac{c_0}{n}, \frac{2c_0}{n} - (\frac{c_0}{n})^2]$ where c_0 is

the cutoff constant for paths of length $d \log n$, $d \geq 1$. This provides a partial answer to the problem of finding the depth of $G_{n,p}$ for $p = \theta(\frac{1}{n})$, open up to now. The remaining cases of $p = o(\frac{1}{n})$ and $p = \Omega(\frac{1}{n})$ can be derived by results of [Erdös, Renyi, 59] (see 2.2 and 2.3 of the text).

Consider the following two-stage experiment: Draw an instance of G_{n,p_1} with $p_1 = \frac{c'}{n}$ $(c' > 0)$. Then, for each edge which failed to appear, draw again with probability $p_2 = \frac{c'}{n}$. The result of the combined experiments is an instance of the graph $G_{n,p}$ with $p = p_1 + (1 - p_1)p_2 = \frac{2c'}{n} - (\frac{c'}{n})^2 = \theta(\frac{c'}{n})$. The first stage of the above experiment constructs an instance I of G_{n,p_1}. Let E be the event "the instance I has a path of length $\delta \cdot \log n$ (where δ is a fixed constant)" and $\overline{E}$ be the complement of E. Let us consider the probability of E. We shall need the following Lemma:

LEMMA 2.1 ([Erdös, Renyi, 59]): Given $G_{n,p}$ with $p = \frac{c}{n}$, let $\tau_{\delta \cdot \log n}$ be the number of paths of length $\delta \cdot \log n$ in $G_{n,p}$. Then

$$\lim_{n \to \infty} \text{Prob}\,\{\tau_{\delta \cdot \log n} = 0\} = e^{-\lambda}$$

with

$$\lambda = (2c)^{\delta \cdot \log n - 1} \frac{(\delta \cdot \log n)^{\delta \cdot \log n - 2}}{(\delta \cdot \log n)!} .$$

The above lemma can be used to prove that " $\exists c_0 > 0$: for $c \geq c_0$ the graph $G_{n,p}$ with $p = \frac{c}{n}$ has at least one path of length $\delta \cdot \log n$ with probability $\rightarrow 1$ as $n \rightarrow \infty$ and for $c < c_0$ the $G_{n,p}$ with $p = \frac{c}{n}$ has no such path with probability $\rightarrow 1$ as $n \rightarrow \infty$" (to prove that, find the c_0 such that $\lambda(c) \rightarrow \infty$ if $c \geq c_0$ and $\lambda(c) \rightarrow 0$ if $c < c_0$ for $n \rightarrow \infty$). Assume in our construction that $c' \geq c_0$. Then, Prob $(E) \geq 1 - n^{-\alpha'(c')}$, where $\alpha'(c') > 2$ is a constant depending on c'. Let us condition the following on the event E.

We start from the vertices of the path T of length $\delta \cdot \log n$ (at least such a path exists, by E) and use them as the initial border set B_0 for a breath-first-search (BFS) process in the graph induced on the vertices of $G_{n,p}$ by just the second stage of the experiment.

Let $|B_0| = \delta \cdot \log n$ and let us find the distribution of the number of vertices to which the vertices of T are immediate neighbors. The set of these vertices will be the new border set of the second stage of BFS, and let us call it B_1. We are interested in the Prob $\{|B_1| \geq 2 \cdot \delta \cdot \log n\}$. Let $|S_0|$ be equal to $|B_0| = \delta \cdot \log n$ (the number of visited vertices at the beginning of the first stage of BFS). The unvisited vertices are $n - |S_0| = n - \delta \cdot \log n$ and each of them has probability $h_0 = 1 - (1 - p_2)^{\delta \cdot \log n}$ to be connected to B_0. Hence

$$\text{Prob } \{|B_1| = x\} = \binom{n-|S_0|}{x} h_0^x (1 - h_0)^{n-\delta \log n - x} .$$

But $h_0 = 1 - \left(1 - \frac{c'}{n}\right)^{\delta \cdot \log n}$ and since $\frac{\log n}{n} \rightarrow 0$ as $n \rightarrow \infty$ we have that $h_0 \rightarrow \frac{c'\delta \log n}{n}$ as $n \rightarrow \infty$. The mean number of $|B_1|$ is

$$(n - |S_0|)h_0 = c'\delta \log n - \frac{c'\delta^2 \log^2 n}{n} \to c'\delta \log n \quad \text{as } n \to \infty .$$

Let $c' > 2$ and let us find the probability Prob $\{|B_1| < 2\delta \cdot \log n = 2|B_0|\}$. From [Feller, 57] pg. 140 this is less than

$$Q = \binom{n - S_0}{2\delta \log n} h_0^{2\delta \log n} (1 - h_0)^{n - \delta \log n - 2\delta \log n} \cdot r$$

where

$$r = \frac{(n - \delta \log n - 2\delta \log n + 1)h_0}{(n - \delta \log n + 1)h_0 - 2\delta \log n} = \frac{c'\delta \log n}{c'\delta \log n - 2\delta \log n} \to \frac{c'}{c' - 2} \quad \text{as } n \to \infty .$$

Then Q is asymptotically equal to

$$\binom{n - \delta \log n}{2\delta \log n} \left(\frac{c' \delta \log n}{n}\right)^{2\delta \log n} \left(1 - \frac{c' \delta \log n}{n}\right)^{n-3\delta \log n} \cdot \frac{c'}{c' - 2}$$

$$= n^{2\delta \log (c'/2)} \cdot n^{-c'\delta} \cdot \frac{c'}{c'-2} < n^{-\delta(c'-2 \log (c'/2))} = n^{-\beta} ,$$

for

$$\beta = \delta\left(c' - 2 \log (c'/2)\right) .$$

Hence, Prob $\{|B_1| \geq 2\cdot |B_0|\} \geq 1 - n^{-\beta}$. We now inductively form sets S_k

and B_k such that $S_k = B_0 + B_1 + \ldots + B_k$ and $h_k = 1 - (1 - p_2)^{n-S_k}$ and B_{k+1} being selected according to the density

$$\text{Prob}\ \{B_{k+1} = x\} = \binom{n - S_k}{x} h_k^x (1 - h_k)^{n-S_k-x} .$$

(This models the BFS process.)

We claim that, as soon as $|S_k| \leq \frac{n}{2}$,

$$\text{Prob}\ \{|B_i| \geq 2 \cdot |B_{i-1}| \text{ given } F\} \geq 1 - n^{-\beta}$$

for

$$F = \bigwedge_{j=1}^{i-1} (|B_j| \geq 2|B_{j-1}|) .$$

This is so, since the first extension (from the smallest border set $B_0 = \delta \cdot \log n$) is the most difficult. Then, the event

$$E_1 = \bigwedge_{i=1}^{k+1} (|B_i| \geq 2|B_{i-1}|)$$

has probability $\geq \prod_{i=1}^{k+1} (1 - n^{-\beta}) \geq 1 - n^{-\beta+1}$ and let k be the last index for which $|S_k| \leq \frac{n}{2}$. Set $|S_k| = \frac{n}{2}(1 - \varepsilon)$ with $|B_k| \geq \frac{|S_k|}{2} = \frac{n}{4}(1 - \varepsilon)$. From $|S_k| = |B_0| + \ldots + |B_k| \geq |B_0| + 2|B_0| + \ldots + 2^k|B_0|$ we get $|S_k| \geq (2^{k+1} - 1)|B_0|$, implying

$$
\begin{aligned}
k &= \log\left(\frac{S_k}{B_0} + 1\right) - 1 \\
&= \log\left(\frac{n}{2}\,\frac{(1-\varepsilon)}{\delta \log n} + 1\right) - 1 \\
&= \log n + \log \frac{1-\varepsilon}{2\delta} - \log\log n - 1 .
\end{aligned}
$$

Hence

$$S_{k+1} = S_k + B_{k+1} \geqslant S_k + 2B_k \geqslant \frac{n}{2}(1-\varepsilon) + \frac{n}{2}(1-\varepsilon) .$$

So, we proved that:

LEMMA 2.2: Conditioned on the event E, *the BFS process will visit at least* $n(1-\varepsilon)$ *nodes of* G_{n,p_2} *in* τ_1 *stages where* $\tau_1 \leqslant \log n$, *with probabiity* $\geqslant 1 - n^{-\beta+1}$.

Let us allow the BFS to continue until it can visit no more vertices, and let f be the final stage (such that $|B_{f+1}| = 0$). Consider the stage $f' \leqslant f$ such that $|S_{f'}| \leqslant n - \log n$ but either $|S_{f'+1}| > n - \log n$ or $|B_{f'+1}| = 0$. After f' either the process stops or there are at most $\delta \cdot \log n$ nodes more to visit and this cannot take more than $\delta \cdot \log n$ stages in either case. At f', the set $B_{f'}$, of "unvisited" nodes is at least of size $\delta \cdot \log n$. Consider the backward process in which the set of unvisited nodes increases as stages are

executed in reverse. The new process is symmetric to the BFS and because of the symmetry we can argue that:

LEMMA 2.3: *With probability at least* $1 - n^{-\beta+1}$, *the process of backward BFS will reach a set of at least* $n(1 - \varepsilon)$ *nodes in number of stages* $\tau_2 \leq \log n$.

By Lemmas 2.2 and 2.3, there is a stage number τ' such that $f' - \tau_2 \leq \tau' \leq \tau_1 \leq f$ in which the border sets of the two processes either coincide or one is a subset of the other. Hence, $\tau' + \tau_1$ is an upper bound to number of stages up to stage f and the total number of stages up to the end of the (forward) BFS is $\tau' + \tau_1 + \delta \log n \leq (2 + \delta) \log n$. The probability of both the events considered in Lemmas 2.2 and 2.3 together is at least $1 - 2n^{-\beta+1}$. Hence we have proved that

LEMMA 2.4: *The diameter of the larger component (and hence the depth of)* $G_{n,p}$ *is* $\leq (2 + \delta) \cdot \log n$ *with (conditional on* $\mathcal{E}$*) probability at least* $1 - 2n^{-\beta+1}$.

Dropping condition $\mathcal{E}$, we get that:

The depth of $G_{n,p}$ with $p \geq \frac{2c'}{n} - \left(\frac{c'}{n}\right)^2$ is $\leq (2 + \delta) \log n$ with

probability at least $1 - \max\{2n^{-\beta+1}, n^{-\alpha(c')}\}$, where $\beta = \delta\big(c' - 2\log(c'/2)\big)$.

By selecting δ and c' suitably, we can guarantee a probability of at least $1 - n^{-\gamma}$, with $\gamma > 2$.

Note that for $p \leq c'/n$. Lemma 2.4 implies Prob $\{\text{depth of } G_{n,p} \leq \delta \cdot \log n\} \to 1$ as $n \to \infty$. Hence, our result holds for any p, $p \not\in [\frac{c'}{n}, \frac{2c'}{n} - (\frac{c'}{n})^2]$, where c' is the cutoff constant for paths of length $\delta \cdot \log n$, $\delta > 2$. This proves Theorem 1.1 with $c_0 = c'$ for the case $p = \theta(\frac{1}{n})$. □

2.2. *The Case of* $p = o(n^{-1})$

LEMMA 2.5 (Erdös, Renyi, 59]): *Let the "average valence" of a graph of* n *vertices and* m *edges be the number* $2n/m$. *Let a graph be called "balanced" if its average valence is greater than, or equal to the average valance of any of its subgraphs. Then, the probability that* $G_{n,p}$ *contains at least one member of the class* $B_{k,\ell}$ *of balanced graphs of* k *vertices and* ℓ *edges is*

$$\text{Prob}\ \{|B_{k,\ell}| \geq 1\} = O\Big(\frac{p}{n^{-k/\ell}}\Big)\ .$$

COROLLARY 2.5: Prob $\{\text{depth of } G_{n,p} < \log n\} \to 1$ *as* $n \to \infty$ *for* $p = o(n^{-1})$.

PROOF: For $p = o(n^{-1})$ we can always write p as $p = (c/n)\cdot\omega^{-1}(n)$ where $c > 0$ is a constant and $\omega(n) \to \infty$ as $n \to \infty$. Let $B = B_{\log n, \log n-1}$ be the class of paths of log n vertices. From Lemma 2.5 we get

Prob $\{G_{n,p}$ contains at least one member of B$\}$

$$= O\left(\frac{c}{n\cdot\omega(n)\cdot n^{-\log n/\log n-1}}\right) = O\left(\frac{c}{\omega(n)}\right) \to 0 \quad \text{as } n \to \infty.$$

□

The above implies that the average depth of $G_{n,p}$ is less than log n for $p = o(n^{-1})$ and all $n > n_0$ for some n_0.

2.3. *The Case of* $p = \Omega(n^{-1})$

LEMMA 2.6 ([Erdös, Renyi, 59]): *Let* $p = (c/n)\cdot\omega(n)$ *where* $c > 0$ *is a constant and* $\omega(n) \to \infty$ *as* $n \to \infty$. *Then, if* $|B_{k,\ell}|$ *is the number of balanced graphs of the class* $B_{k,\ell}$ (k *vertices* ℓ *edges) contained in* $G_{n,p}$ *we have*

$$\text{Prob } \{|B_{k,\ell}| < \omega(n)^{\ell}\} = O\left(\frac{1}{\omega(n)}\right)$$

implying

$$\text{Prob}\ \{|B_{k,\ell}| = 0\} = O\left(\left(\frac{1}{\omega(n)}\right)^{\ell}\right) .$$

COROLLARY 2.6: *The depth of* $G_{n,p}$ *is* $\leq \log n$ *for* $p = \Omega(n^{-1})$, *with probability going to* 1 *as* $n \to \infty$. *This implies that, for* $p = \Omega(n^{-1})$, *the average depth of* $G_{n,p}$ *is not more than* $\log n$ *for all* $n > n_0$, *for some* $n_0 > 0$.

PROOF: Let $B_{k,\ell}$ be the class of binary balanced trees of n vertices, n - 1 edges. Then

$$\text{Prob}\ \{|B_{k,\ell}| \geq 1\} \geq 1 - O\left(\left(\frac{1}{\omega(n)}\right)^{n-1}\right) .$$

Hence

$$\text{Prob}\ \{|B_{k,\ell} \geq 1\} \to 1 \quad \text{as } n \to \infty .$$

□

3. THE CASE OF HIGH DENSITIES, $p \geq cn^{-1/3}$, $c > 1$

LEMMA 3.1: *The depth of $G_{n,p}$, for $p \geq cn^{-1/3}$ is ≤ 3, with probability tending to 1 as n goes to ∞. Here $c > 1$ is a constant.*

PROOF: Consider any pair u, v of vertices $G_{n,p}$. Let S_u (resp. S_v) be the sets of neighbors of u (resp. of v). Then, for any $\beta \in (0,1)$

$$\text{Prob}\ \{|S_u| \geq p(n-1)(1-\beta)\} \geq 1 - e^{-\beta^2(n-1)/2}$$

by the Chernoff bounds [Chernoff, 52]. For $\beta = 1/2$ we get

$$\text{Prob}\ \{|S_u| \geq \frac{cn^{2/3}}{2}\} \geq 1 - e^{-(c/8)n^{2/3}} .$$

Similarly,

$$\text{Prob}\ \{|S_v| \geq \frac{cn^{2/3}}{2}\} \geq 1 - e^{-(c/8)\cdot n^{2/3}} .$$

Hence

$$\text{Prob}\ \{\text{both } |S_u|, |S_v| \geq \frac{cn^{2/3}}{2}\} \geq 1 - 2e^{-(c/8)n^{2/3}} .$$

Let E be the event "both $|S_u|$ $|S_v| \geq \frac{cn^{2/3}}{2}$". Then,

Prob {no edge between S_u, S_v given E} $\leqslant (1-p)^{|S_u|\cdot|S_v|} \leqslant \left(1 - \frac{c}{\sqrt[3]{n}}\right)^{\frac{c^2}{4} n^{4/3}}$

$$\to e^{-(c^3/4)\cdot n} .$$

Hence, combining terms,

Prob {$\exists$ path of length $\leqslant 3$ between u,v}

$$\geqslant \left(1 - 2e^{-(c/8)\cdot n^{2/3}}\right)\left(1 - e^{-(c^3/4)\cdot n}\right) \geqslant 1 - 3e^{-(c/8)\cdot n^{2/3}} .$$

Hence, Prob {$\forall$ pairs u,v $\exists$ path of length $\leqslant 3$ between them}

$$\geqslant 1 - 3n^2 e^{-(c/8)n^{2/3}} \to 1 \quad \text{as } n \to \infty.$$

REFERENCES

Angluin, D. and L.G. Valiant, " Fast probabilistic algorithms for Hamiltonian paths and matchings," J. Comp. Syst Sci. 18 (1979), pp 155-193.

Chernoff, H., "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," Annals of Math. Statistics, Vol. 23, 1952.

Erdös, P. and A. Renyi, "On the evolution of random graphs," Pub. Math. Inst. Hung. Acad. Sci. 5A, 1960, pp. 17-61.

Feller, W., "An introduction to probability theory and its applications," Vol. 1, 3rd edition, John Wiley and Sons, New York, 1968.

Karp, R.M., "The probabilistic analysis of combinatorial search algorithms," Algorithms and Complexity: New Directions and Recent Results, J.F. Traub, Ed. Acad. Press, New York, 1976, pp. 1-19.

Karp, R.M. and M. Sipser, "Maximum matchings in sparse random graphs," 22nd Sym. on Foundations of Computer Science, Nashville, Tenn., Oct. 1981, pp. 364-375.

NYU c.2
Comp. Sci. Dept.
TR-82
Reif
On the depth of a random
graph.

c.2
NYU
Comp. Sci. Dept.
TR-82 Reif

AUTHOR
On the depth of a random

TITLE
graph.

DATE DUE	BORROWER'S NAME

This book may be kept NOV 1 1983

FOURTEEN DAYS

A fine will be charged for each day the book is kept overtime.

GAYLORD 142			PRINTED IN U S A